Calhoun: The NPS Institutional Archive

DSpace Repository

Theses and Dissertations 1. Thesis and Dissertation Collection, all items

1971

# A modified Kolomogorov-Smirnov test applicable to censored samples

Davidson, Teddy George

Monterey, California; Naval Postgraduate School

http://hdl.handle.net/10945/15816

This publication is a work of the U.S. Government as defined in Title 17, United States Code, Section 101. Copyright protection is not available for this work in the United States.

*Downloaded from NPS Archive: Calhoun*

Calhoun is the Naval Postgraduate School's public access digital repository for research materials and institutional publications created by the NPS community. Calhoun is named for Professor of Mathematics Guy K. Calhoun, NPS's first appointed -- and published -- scholarly author.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

http://www.nps.edu/library

# A MODIFIED KOLOMOGOROV-SMIRNOV TEST APPLICABLE TO CENSORED SAMPLES

TEDDY GEORGE DAVIDSON

LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIF. 93940

United States
Naval Postgraduate School

LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIF. 93940

# THESIS

## A MODIFIED KOLOMOGOROV-SMIRNOV TEST APPLICABLE TO CENSORED SAMPLES

by

Teddy George Davidson

Thesis Advisor: D. R. Barr

September 1971

*Approved for public release; distribution unlimited.*

# A Modified Kolomogorov-Smirnov Test
# Applicable to Censored Samples

by

Teddy George Davidson
Lieutenant, United States Navy
B.S., University of Washington, 1964

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1971

# ABSTRACT

Herein is presented a derivation and computational formulation for a modified Kolomogorov-Smirnov test. This test extends the hypothesis testing and confidence limits advantages of the Kolomogorov-Smirnov test to data which is censored beyond a predetermined number of observations.

A listing of the computer program used in the calculations of significance levels and the resulting significance levels for specified parameter values for the modified test are included.

## TABLE OF CONTENTS

I. INTRODUCTION ---------------------------------- 4

II. KOLOMOGOROV-SMIRNOV TEST ------------------------ 6

III. EXISTENCE OF THE MODIFIED KOLOMOGOROV-SMIRNOV TEST -------------------------------------------- 8

IV. PROOF THAT MODIFIED KOLOMOGOROV-SMIRNOV TEST IS DISTRIBUTION FREE ------------------------------- 10

V. DERIVATION OF CRITICAL VALUES ($D_{f,n}$) ----------- 12

VI. COMPUTATIONAL FORMULATION FOR CALCULATION OF SIGNIFICANCE LEVELS ------------------------------ 19

VII. VERIFICATION OF SIGNIFICANCE LEVEL COMPUTATIONS - 26

VIII. APPLICATIONS OF MODIFIED KOLOMOGOROV-SMIRNOV TEST -------------------------------------------- 29

IX. WORK REMAINING TO BE DONE ---------------------- 30

X. SUMMARY ---------------------------------------- 31

APPENDIX A: COMPUTER PROGRAM FOR CALCULATION OF SIGNIFICANCE LEVELS --------------------------- 32

APPENDIX B: CALCULATED SIGNIFICANCE LEVELS ----------- 33

APPENDIX C: COMPUTER PROGRAM FOR VERIFICATION OF CALCULATED SIGNIFICANCE LEVELS -------------- 40

BIBLIOGRAPHY ---------------------------------------- 41

INITIAL DISTRIBUTION LIST -------------------------- 42

FORM DD 1473 ---------------------------------------- 43

# I. INTRODUCTION

In statistical applications, it is frequently desired to determine, with some specified confidence, whether a random sample is from a hypothesized distribution. Similarly, it is frequently desirable to establish confidence limits on the distribution function using sample observations. There are numerous procedures which accomplish these objectives when the sample is uncensored. The Kolomogorov-Smirnov test is applicable to both of these objectives and has the advantages of being distribution free and simple to apply.

In situations wherein the data is censored beyond a predetermined number of failures, the complete sample is not available since the values associated with sample observations greater than the censoring point are unobservable. In such a situation, one still might desire to conduct a "Kolomogorov-Smirnov type" test over only the ordered uncensored sample observations. Such a modified Kolomogorov-Smirnov test is derived herein. To the knowledge of the author, this test has not been previously available.

The modified Kolomogorov-Smirnov test is proved to be distribution free. Furthermore, a computational formulation is presented for calculating significance levels and the significance levels are tabulated as a function of sample size and critical value.

Proofs and derivations contained herein do not specifically deal with the usual problems with events of probability zero, however they can readily be modified to include such events. All computer work performed in connection with this paper was conducted at the United States Naval Postgraduate School on an IBM 360 computer.

The following listing constitutes the description and/or definition of notation utilized herein:

| Notation | Description |
|---|---|
| $S_n$ | Cumulative distribution function of a random sample of size n. |
| $F$ | Hypothesized distribution function. |
| $f$ | Number of observations over which the modified test is run. |
| $\alpha$ | Level of test. |
| $n$ | Sample size. |
| $d, D$ | Critical values. |
| $D_{f,n}$ | Critical value associated with a test run over the first f ordered observation from sample size n. |
| $x_1 < x_2 < -- < x_n$ | Ordered observed random sample from distribution F. |

$$P_{f,n}(D) = P\left( \sup_{x \le x_f} \left| S_n(x) - F(x) \right| < D_{f,n} \right)$$

## II. KOLOMOGOROV-SMIRNOV TEST

Kolomogorov established that the distribution of the statistic

$$\sup_x \mid S_n(x) - F(x) \mid$$

is independent of the parent population distribution F if F is continuous, and he derived its limiting distribution. Massey [3] and Miller [4] present methods by which significance level $\alpha$ can be computed for a given critical value d such that

$$P( \sup_x \mid S_n(x) - F(x) \mid \geq d ) = \alpha.$$

This relation between significance level versus critical value and sample size is the basis for applying the Kolomogorov-Smirnov test. Further information on the derivation, tabulation and use of the test is contained in References [1], [2], [3] and [4].

The essence of the Kolomogorov-Smirnov test is that, for a sample distribution function ($S_n$) and a continuous parent population distribution function (F), the statistic

$$\sup \mid S_n - F \mid$$

is independent of the distribution of the parent population. Thus, given that the population has distribution F, the probability that the maximum absolute difference between F and the sample distribution will exceed a specified critical value is dependent only upon the sample size and

the critical value specified. The test itself then is simple to conduct since it entails only the determination of this maximum difference and whether it is less than the specified critical value $D_n$. Furthermore, as stated by Kendall and Stuart (1, pg 457) "we may reverse the procedure of testing for fit and use $D_n$ to set confidence limits for a (continuous) distribution function as a whole."

The Kolomogorov-Smirnov test is thus a simple method for either testing goodness-of-fit or establishing confidence limits for a distribution function from a sample when a continuous parent population distribution is hypothesized.

## III. EXISTENCE OF THE MODIFIED KOLOMOGOROV-SMIRNOV TEST

The Kolomogorov-Smirnov goodness of fit test establishes a band $S_n \pm d$ such that the probability is $(1-\alpha)$ that the true distribution function F lies entirely within this band, [1]. That is,

$$P( \sup_{x} | S_n(x) - F(x) | > d ) = \alpha,$$

where $\alpha$ is the probability that the test rejects the hypothesis that F is the distribution of X, given F is the distribution of X and where d is the critical value for test size $\alpha$. Let this test be denoted by T. Then T rejects iff $\sup_{x} | S_n(x) - F(x) | > d$ for some x.

The modified Kolomogorov-Smirnov test proposed herein is to apply the Kolomogorov-Smirnov test using d* for critical value and applying the test only to the first f order statistics to yield a significance level $\alpha$. That is,

$$P( \sup_{x \leq x_f} | S_n(x) - F(x) | > d^* ) = \alpha.$$

Let this test be denoted by T*.

Let A denote the event that rejection occurs for some $x_i$, $i \leq f$ when using test T, given that test T rejects. Then, with each test using critical value d*,

$$\begin{aligned} P(T^* \text{ rejects}) &= P(T \text{ rejects})\ P(\text{rejection is in first f order statistics/rejection occurs}) \\ &= P(T \text{ rejects})\ P(A) \leq P(T \text{ rejects}) = \alpha. \end{aligned}$$

However, P(T rejects) is a continuous monotone decreasing function of d. Therefore, for each d there is a d* ($d^* \leq d$) such that

$$\begin{aligned} P(T \text{ rejects using } d) &= P(T \text{ rejects using } d^*)\, P(A) \\ &= P(T^* \text{ rejects using } d^*) \end{aligned}$$

which establishes the existence of the test T*.

## IV. PROOF THAT THE MODIFIED KOLOMOGOROV-SMIRNOV TEST IS DISTRIBUTION FREE

Let $Y_1, Y_2, \ldots, Y_n$ be an ordered random sample from a continuous distribution F. Define

$$W = \sup_{y \leq Y_f} | S_n(y) - F(y) |$$

where $n \cdot S_n(y) = \max \{k: \ Y_k \leq y\}$. Let

$X_i = F(Y_i)$ which may be written "$X_i(Y)$"

and

$$W^* = \sup_{x \leq X_f} | S_n^*(x) - x | .$$

Then $X_1, X_2, \ldots, X_n$ is (almost surely) an ordered sample from the uniform distribution over the interval (0,1).

However,

$$\begin{aligned} W &= \sup_{y \leq Y_f} | S_n(y) - F(y) | \\ &= \sup_{y \leq Y_f} \frac{1}{n} | \max \{k: \ F(Y_k) \leq F(y)\} - n\, F(y) | \\ &= \sup_{\{y: F(y) \leq F(Y_f)\}} \frac{1}{n} | \max \{k: F(Y_k) \leq F(y)\} - n\, F(y) | \\ &= \sup_{\{y: X(y) \leq X_f\}} \frac{1}{n} | \max \{k: X_k \leq X(y)\} - n\, X(y) | \\ &= \sup_{x \leq X_f} \frac{1}{n} | \max \{k: X_k \leq x\} - n\, x | \\ &= \sup_{x \leq X_f} | S_n^*(x) - x | \\ &= W^* . \end{aligned}$$

Thus $W = W^*$ and it follows that for any critical value D

$$P\,(W \geq D) = P\,(W^* \geq D).$$

Since $W^*$ has a distribution not depending upon F, it follows that the test is distribution free.

## V. DERIVATION OF CRITICAL VALUES ($D_{f,n}$)

The significance level, $P_{f,n}(D)$, associated with a given critical value, $D_{f,n}$, can be calculated by the following procedure. This procedure was motivated by, and is an analytical generalization of the procedure presented by Massey [3] to calculate Kolomogorov-Smirnov critical values.

Since it has been shown in Section IV that $P_{f,n}(D)$ is independent of F provided only that F is continuous, it is sufficient to consider only the case

$$F(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ if } 0 \leq x \leq 1 \\ 1 \text{ if } x > 1. \end{cases}$$

Divide the interval (0,1) into n parts by the points 1/n, 2/n,---,(n-1)/n. The step function $S_n$ rises by jumps of exactly 1/n. Thus, in order to be inside the band $F(x) \pm k/n$ at $x = i/n$, $S_n(i/n)$ must be one of the lattice points whose ordinates are $(i-k+1)/n$, $(i-k+2)/n$,---,$(i+k-1)/n$. Let $\alpha_i$ be an n dimensional vector of non-negative integers

$$\alpha_i = (\alpha_{1,i}, \alpha_{2,i}, \alpha_{3,i}, ---, \alpha_{n,i})$$

such that

$$\sum_{j=1}^{n} \alpha_{j,i} = n.$$

Suppose that the step function $S_n$ stays inside the band by means of $\alpha_{j,i}$ of the observations from F falling within the

interval $((j-1)/n,\ j/n)$, $j = 1, 2, \cdots, n$. Under $H_0$, the probability of this happening is given by the multinomial law as

$$p(\alpha_{1,i}, \cdots, \alpha_{n,i}) = \frac{n!}{\alpha_{1,i}!,\ \alpha_{2,i}! \cdots \alpha_{n,i}!} \cdot \left(\frac{1}{n}\right)^{\sum_{j=1}^{n} \alpha_{j,i}}$$

$$= \frac{1}{\prod_{j=1}^{n} \alpha_{j,i}!} \cdot \frac{n!}{n^n} \quad .$$

Thus the probability of the step function staying within the band F ± $D_{f,n}$ is given by

$$P_{f,n}(D) = \sum_i \frac{1}{\prod_{j=1}^{n} \alpha_{j,i}!} \quad \frac{n!}{n^n} = \frac{n!}{n^n} \sum_i \frac{1}{\prod_{j=1}^{n} \alpha_{j,i}!}$$

where the summation is over all i for which $\alpha_i$ is associated with $S_n$ within the band. There exists a one-to-one correspondence between $\{\alpha_i\}$ and the paths of $S_n$.

Define $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_k)$, where the elements of $\alpha$ represent all paths of $S_n$. Each element of $\alpha$ represents either a path for which $S_n(x)$ remains entirely within the band for $x \leq x_f$ or a path for which $S_n(x)$ does not remain entirely within the band for $x \leq x_f$.

Let $I = \{\alpha_i;\ S_n(x)$ remains entirely within the band for $x \leq x_f\}$. Then by the multinomial law

$$P\left(\sup_{x \leq x_f} | S_n(x) - x | < D_{f,n}\right) = \frac{n!}{n^n} \sum_{i \varepsilon I} \frac{1}{\prod_{j=1}^{n} \alpha_{j,i}!} \ .$$

Consider all possible values of $\alpha_{1,k}$ (that is $\alpha_{1,k} = 0, 1, \cdots, n$). Then either

(a) $\alpha_{1,k} \geq f$

or

(b) $\alpha_{1,k} < f$

and one of the following mutually exclusive cases must hold (refer to Figure 1).

A11. if $\alpha_{1,k} \geq f$ and $S_n(x)$ is entirely within the band for $x \leq x_f$, then $\alpha_k \varepsilon I$ regardless of the values of $\alpha_{2,k}$ --- $\alpha_{n,k}$. The contribution of $\alpha_k$ to $P_{f,n}(D)$ is

$$\frac{n!}{n^n} \frac{1}{\alpha_{1,k}!} \sum \frac{1}{\alpha_{2,k}! \; --- \; \alpha_{n,k}!}$$

where the sum is over all $\alpha_{j,k}$ such that $\sum_{j=2}^{n} \alpha_{j,k} = n - \alpha_{1,k}$.

A21. if $\alpha_{1,k} \geq f$ and $S_n(x)$ is not entirely within the band for $x \leq x_f$, then $\alpha_k \not\varepsilon I$ regardless of the value of $\alpha_{2,k}$, ---, $\alpha_{n,k}$ and therefore $\alpha_k$ does not contribute to the sum which composes $P_{f,n}(D)$.

B11. if $\alpha_{1,k} < f$ and $S_n(x)$ is entirely within the band for $x \leq x_{\alpha_{1,k}}$, then whether $\alpha_k \varepsilon I$ will depend upon $\alpha_{2,k}$, ---, $\alpha_{n,k}$ and the contribution of $\alpha_k$ to $P_{f,n}(D)$ is dependent upon $\alpha_{2,k}$, $\alpha_{3,k}$, ---, $\alpha_{n,k}$.

B21. if $\alpha_{1,k} < f$ and $S_n(x)$ is not entirely within the band for $x \leq x_{\alpha_{1,k}}$, then $\alpha_k \not\varepsilon I$ regardless of $\alpha_{2,k}$, $\alpha_{3,k}$, ---, $\alpha_{n,k}$ and therefore $\alpha_k$ does not contribute to the sum which composes $P_{f,n}(D)$.

Consider next $\alpha_k = (\alpha_{1,k}, \alpha_{2,k}, ---, \alpha_{n,k})$ such that $\alpha_{1,k} \Longrightarrow$ B11. Then either

$$\text{(a) } \alpha_{1,k} + \alpha_{2,k} \geq f$$

or

$$\text{(b) } \alpha_{1,k} + \alpha_{2,k} < f$$

and one of the following mutually exclusive cases must hold (refer to Figure 1).

A12. if $\alpha_{1,k} + \alpha_{2,k} \geq f$ and $S_n(x)$ is entirely within the band for $x \leq x_f$, then $\alpha_k \varepsilon I$ regardless of the values of $\alpha_{3,k}, \alpha_{4,k}, ---, \alpha_{n,k}$ and the contribution of $\alpha_k$ to $P_{f,n}(D)$ is

$$\frac{n!}{n^n} \frac{1}{\alpha_{1,k}! \; \alpha_{2,k}!} \sum \frac{1}{\alpha_{3,k}! \; --- \; \alpha_{n,k}}$$

where the sum is over all $\alpha_{j,k}$ such that $\sum_{j=3}^{n} \alpha_{j,k} = n - \alpha_{1,k} - \alpha_{2,k}$.

A22. if $\alpha_{1,k} + \alpha_{2,k} \geq f$ and $S_n(x)$ is not entirely within the band for $x \leq x_f$, then $\alpha_k \not\varepsilon I$ regardless of $\alpha_{3,k}, ---, \alpha_{n,k}$ and $\alpha_k$ does not contribute to the sum which composes $P_{f,n}(D)$.

B12. if $\alpha_{1,k} + \alpha_{2,k} < f$ and $S_n(x)$ is entirely within the band for $x \leq x_{\alpha_{1,k} + \alpha_{2,k}}$, then whether $\alpha_k \varepsilon I$ will depend upon $\alpha_{3,k}, \alpha_{4,k}, ---, \alpha_{n,k}$ and the contribution of $\alpha_k$ to $P_{f,n}(D)$ is dependent upon $\alpha_{3,k}, ---, \alpha_{n,k}$.

B22. if $\alpha_{1,k} + \alpha_{2,k} < f$ and $S_n(x)$ is not entirely within the band for $x \le x_{\alpha_{1,k} + \alpha_{2,k}}$, then $\alpha_k \notin I$ and $\alpha_k$ does not contribute to the sum which composes $P_{f,n}(D)$.

Consider next $\alpha_k = (\alpha_{1,k}, \alpha_{2,k}, \cdots, \alpha_{n,k})$ such that $\alpha_{1,k} \Rightarrow$ B11 and $\alpha_{2,k} \Rightarrow$ B12. Continue this procedure until the contribution from each $\alpha_i$, $i \varepsilon I$, to $P_{f,n}(D)$ is obtained.

Then

(1) $P_{f,n}(D)$ = the sum over all $i \varepsilon I$ of the contributions to $P_{f,n}(D)$ from $\alpha_i$

$$= \frac{n!}{n^n}\left[ \sum_{\alpha_{1,k} \Rightarrow A11} \frac{1}{\alpha_{1,j}!} \sum \frac{1}{\alpha_{2,j}! \cdots \alpha_{n,j}!} + \sum_{\alpha_{2,k} \Rightarrow A12} \frac{1}{\alpha_{1,j}!\alpha_{2,j}!} \sum \frac{1}{\alpha_{3,j}! \cdots \alpha_{n,j}!} + \cdots \sum_{\alpha_{n-2,j} \Rightarrow A1,n-2} \frac{1}{\alpha_{1,j}! \cdots \alpha_{n-2,j}!} \cdot \frac{1}{\alpha_{n-1,j}!\alpha_{n,j}!} + \sum_{\alpha_{n-1,j} \Rightarrow A1,n-1} \cdot \frac{1}{\alpha_{1,j}! \cdots \alpha_{n,j}!} \right].$$

where the inner sums are over all $\alpha_{i,j}$ such that $\sum_{i=1}^{n} \alpha_{i,j} = n$.

Let $U(k,r) = \sum \frac{1}{\alpha_1! \cdots \alpha_k!}$ over $\alpha$ such that $\sum_{i=1}^{k} \alpha_i = r$.

Then $U(k,r)$ satisfies the difference equation

$$(2) \quad U(m+1,j) = \sum_{h=0}^{j} \frac{1}{(j-h)!} U(m,h) \quad , \ j = 0,\ 1,\ ---\ ,\ r$$

with initial conditions of

$$U(0,0) = 1$$

and

$$U(0,j) = 0 \text{ if } j \neq 0.$$

Thus, by substituting (2) into (1), $P_{f,n}(D)$ is of the form

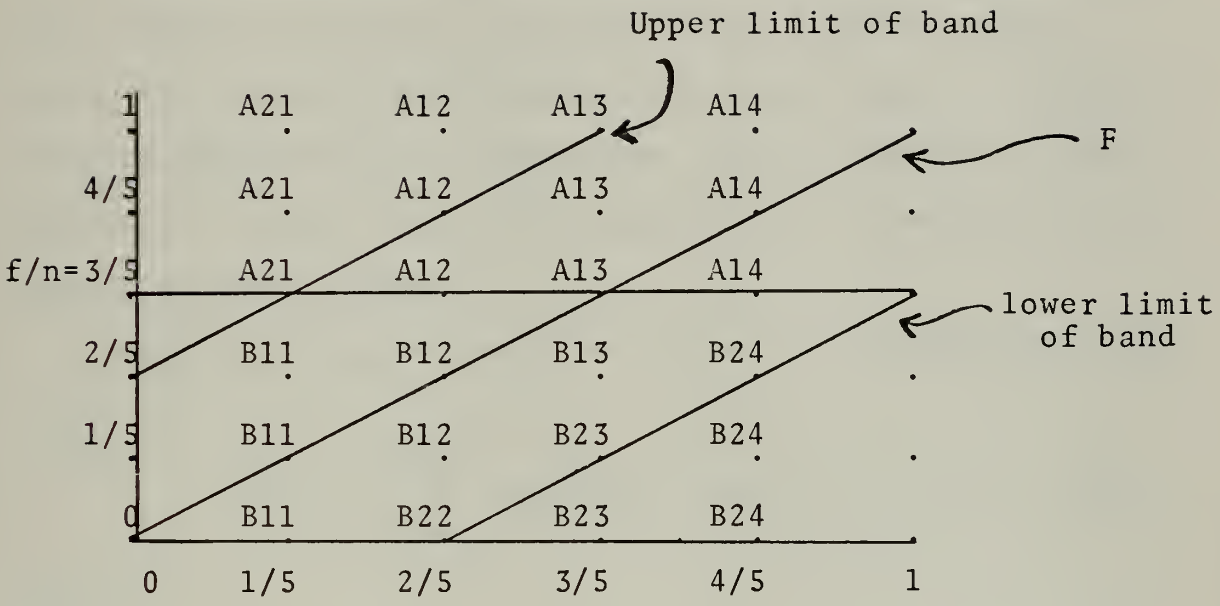
$$P_{f,n}(D) = \frac{n!}{n^n} \Bigg[ \sum_{\alpha_{1,j} \Rrightarrow A11} \frac{1}{\alpha_{1,j}!} U(n-1,\ n-\alpha_{1,j})$$

$$+ \sum_{\alpha_{2,j} \Rrightarrow A12} \frac{1}{\alpha_{1,j}!\ \alpha_{2,j}!} U(n-2,\ n-\alpha_{1,j}-\alpha_{2,j})$$

$$+ - - - + \sum_{\alpha_{n-2,j}} \frac{1}{\alpha_{1,j}!\ ---\ \alpha_{n-2}!} \cdot$$

$$U(2, n-\alpha_{i,j}-\alpha_{2,j} - --- -\alpha_{n-2,j})$$

$$+ \sum_{\alpha_{n-1,j} \Rrightarrow A1,n-1} \frac{1}{\alpha_{1,j}! --- \alpha_{n,j}!} \Bigg].$$



Parameters:

$n = 5$

$D = 2/5$

$f = 3.$

Figure 1. Classification of Lattice Points.

## VI. COMPUTATIONAL FORMULATION FOR CALCULATION OF SIGNIFICANCE LEVELS

The foregoing derivation can be utilized to calculate the significance level associated with a given critical value for the modified Kolomogorov-Smirnov test. However, the procedure would be cumbersome and not amenable to computation. A more tractible computational extension is discussed in what follows.

Massey [3] has derived the set of 2k-1 difference equations

$$U_j(m+1) = \sum_{h=1}^{j+1} \frac{1}{(j+1-h)!} U_h(m) \quad , \; j = 1,2,\text{---},2k-1$$

with

$$U_h(m) = 0 \text{ if } h \geq m + k$$

$$U_i(0) = 0 \text{ for } i \neq k$$

$$U_k(0) = 1 \text{ for } i = k$$

which yields

$$P_n(k/\sqrt{n}\ ) = \frac{n!}{n^n} U_k(n)$$

for the probability that the sample distribution function $(S_n)$ stays entirely within the band F ± k/n.

Referring to the (0,1) interval divided into n equal parts and indexed with 1 through n, the $U_j(m)$ values represent values assigned to the lattice points within the band F ± k/n (refer to Figure 2).

The index m refers to the ordinate index f the lattice point. Thus, the values for the lattice points with

## VI. COMPUTATIONAL FORMULATION FOR CALCULATION OF SIGNIFICANCE LEVELS

The foregoing derivation can be utilized to calculate the significance level associated with a given critical value for the modified Kolomogorov-Smirnov test. However, the procedure would be cumbersome and not amenable to computation. A more tractible computational extension is discussed in what follows.

Massey [3] has derived the set of 2k-1 difference equations

$$U_j(m+1) = \sum_{h=1}^{j+1} \frac{1}{(j+1-h)!} U_h(m) \quad , \; j = 1,2,---,2k-1$$

with

$$U_h(m) = 0 \text{ if } h \geq m + k$$

$$U_i(0) = 0 \text{ for } i \neq k$$

$$U_k(0) = 1 \text{ for } i = k$$

which yields

$$P_n(k/\sqrt{n}\,) = \frac{n!}{n^n} U_k(n)$$

for the probability that the sample distribution function $(S_n)$ stays entirely within the band F ± k/n.

Referring to the (0,1) interval divided into n equal parts and indexed with 1 through n, the $U_j(m)$ values represent values assigned to the lattice points within the band F ± k/n (refer to Figure 2).

The index m refers to the ordinate index of the lattice point. Thus, the values for the lattice points with

ordinate (m+1) depend only upon the values of lattice points within the band with ordinate m and upon the critical value (k/n) and sample size (n).

$U_k(n)$ then represents the value associated with paths of $S_n$ which remain entirely within the band F ± k/n and pass through lattice point (n,n). Consequently $U_k(n)\cdot(n!/n^n)$ is the probability that $S_n$ will remain entirely within the band F ± k/n.

Massey's approach can be modified to provide similar difference equations for the calculation of the significance level associated with critical value (k/n), number of uncensored failures (f), and sample size (n) for the modified Kolomogorov-Smirnov test. The required modification is to replace the requirement that "$S_n$ remain entirely within F ± k/n" with the requirement that "$S_n$ remain entirely within F ± k/n through the $f^{th}$ ordered observation."

The set of difference equations which results from this modification is

$$U_j(m+1) = \sum_{h=1}^{r} \frac{1}{(j+1-h)!} U_h(m) \quad ; \begin{array}{l} m = 0, ---, n \\ j = 0, ---, n \end{array}$$

where the upper limit on the sum is given by

$$r = \begin{cases} n-m+k-1 & \text{if } f < F \pm k/n \\ j+1 & \text{if } f \geq F \pm k/n \text{ (refer to Figure 3).} \end{cases}$$

Additional computational efficiency can be obtained by recalling that the $U_j(m)$ values for each lattice point results from the $\alpha_i$ components which determine specific paths

for $S_n$. Thus, if $\sum_{i=1}^{m} \alpha_i \geq f$ then the first f ordered observations remain within the band F ± k/n regardless of $\alpha_i$, $i = m+1, ---, n$. Consequently, the $U_j(m)$ values for these lattice points can be relabled as $Z_j(m)$ and the difference equations summed only over the lattice points which represent paths for which it is undetermined whether $S_n$ remains within F ± k/n for the first f ordered observations. Referencing Figure 4, it may be seen that this yields the equations

$$U_j(m+1) = \sum_{h \varepsilon C_j} \frac{1}{(j+1-h)!} U_h(m)$$

$$Z_j(m+1) = \sum_{h \varepsilon C_j} \frac{1}{(j+1-h)!} U_h(m)$$

where $C_j$ represents the circled lattice points in Figure 4 having abscissa values less than or equal to j.

Consider the set of difference equations

$$\text{OVUP}(i+1,j) = \sum_{h=0}^{j} \frac{1}{(j+1-h)!} \text{OVUP}(i,j) \quad ; \; j=0,1,---$$

with initial conditions

$$\text{OVUP}(0,0) = 1$$

$$\text{OVUP}(0,j) = 0 \text{ if } j \neq 0.$$

OVUP(i,j) is an array which represents values associated with path segments for which $S_n$ traverses i lattice points to the right while increasing in abscissa value by j lattice points. Consequently, the $Z_j(m)$ values can be translated to

values at lattice point (n,n) by multiplying each $Z_j(m)$ value by OVUP(n-m, n-j-m+f-1). The value of $U_k(n)$ is then the sum of these translated $Z_j(m)$ values. That is

$$U_k(n) = \sum_{\text{all } Z} Z_j(m) \; \text{OVUP}(n-m,\ n-j-m+f-1)$$

and

$$P(F-k/n < S_n < F+k/n) = U_k(n) \frac{n!}{n^n}$$

yields the significance level associated with the critical value k/n, k an integer, for the modified Kolomogorov-Smirnov test.

This computational procedure is specifically applicable to critical values of k/n, k an integer. However, the procedure can readily be adapted to critical values of k/(m·n) with k and m integers by dividing the (0,1) interval into m·n parts vice n parts. Moreover, the significance level associated with critical values not specifically calculated can be approximated by interpolation between critical values for which the significance levels were calculated.

This latter compuational formulation is the basis of the computer program presented as Appendix A. The significance levels as calculated from the program in Appendix A are presented as Appendix B for all combinations of the following parameters:

n = 5, 10, 15, 20, 25, 30

k = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

f = 1, 2, ---, n.

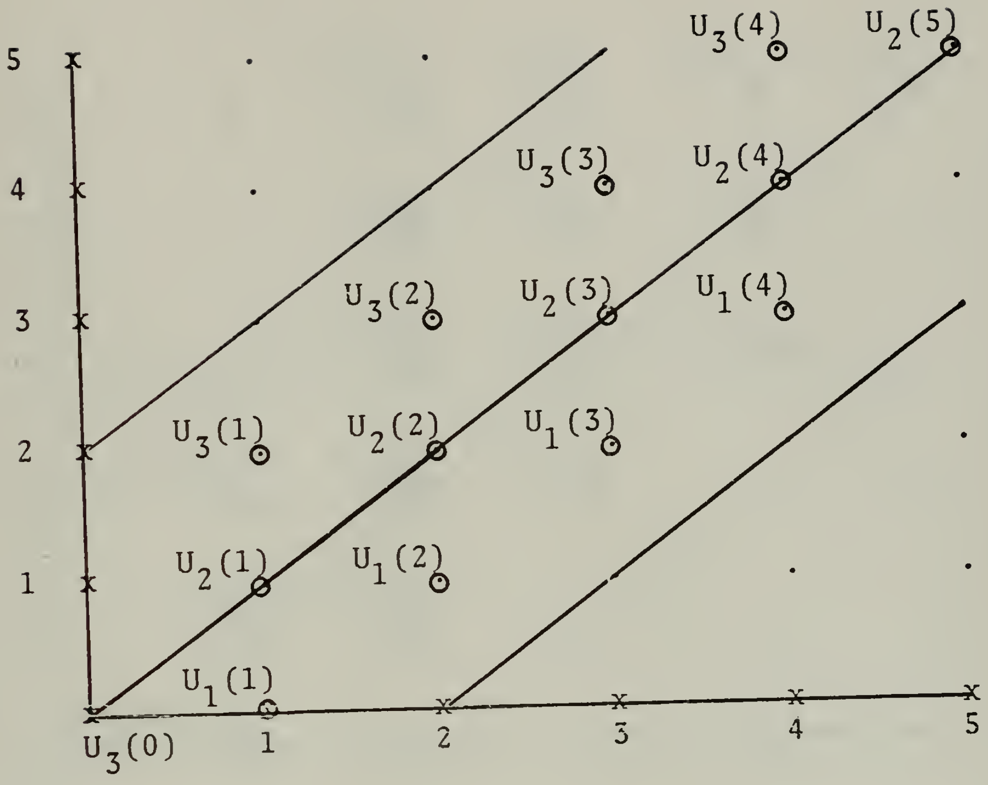


Figure 2. Massey's Procedure.

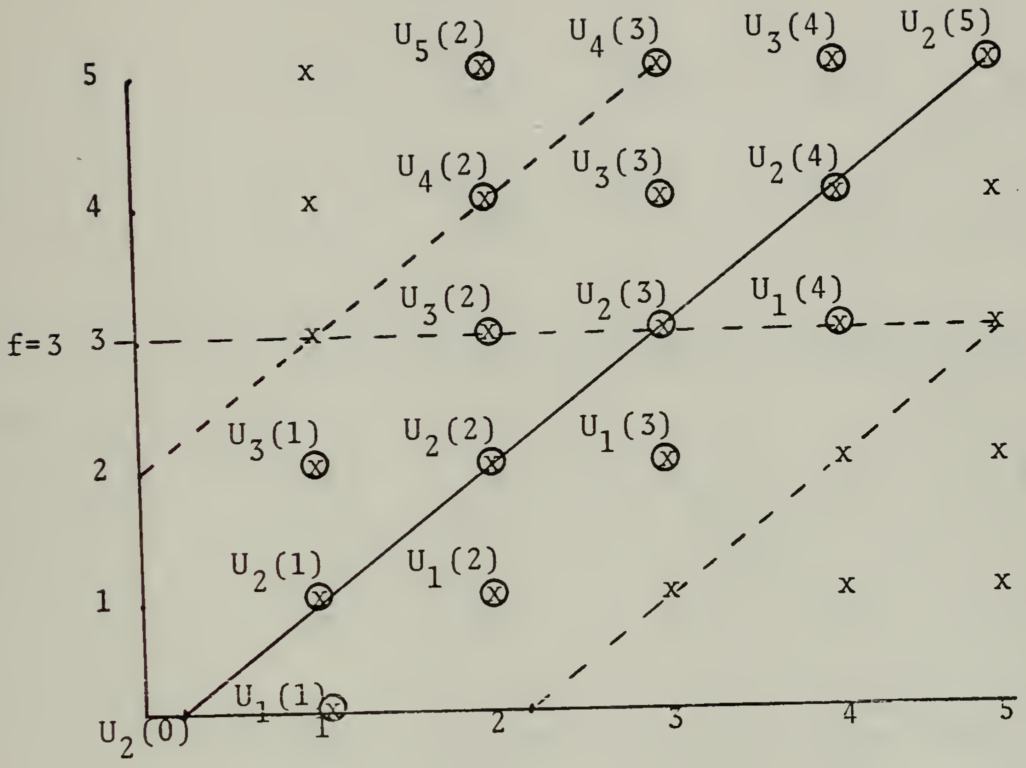


Figure 3. Modified Procedure.

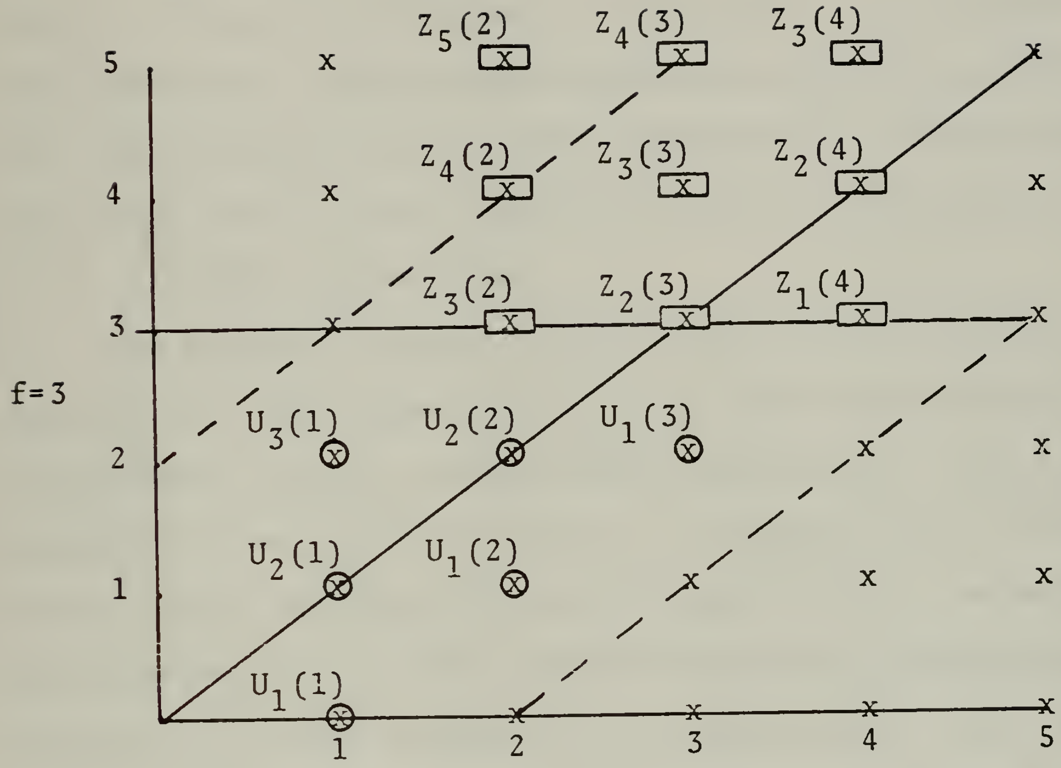


Figure 4. Modified Procedure Using Translation.

## VII. VERIFICATION OF SIGNIFICANCE LEVEL COMPUTATIONS

In order to verify the correctness of the computer program, a computer simulation was conducted. This simulation was based upon a population distribution which was uniform over the interval (0,1). The simulation was run for sample sizes (n) of 5 and 10, with critical value (k/n) of 0.40, and 3000 replications.

The computer program for the simulation is listed in Appendix C. The significance levels resulting from the simulations and the corresponding significance levels from Appendix B are shown in Table I. It appears that the simulated values and the calculated values are in agreement. This contention that the computed values are (approximately) correct is further supported by comparing them with the values calculated by Massey [3]. These values are shown in Table II.

In view of the foregoing, it is concluded that the calculated significance levels (Appendix B) and the computer program from which they were calculated (Appendix A) are accurate and contain no significant errors.

TABLE I

| CRITICAL VALUE | n | # OBSERVATIONS (f) | SIGNIFICANCE LEVEL | |
|---|---|---|---|---|
| | | | SIMULATED | CALCULATED |
| .40 | 5 | 1 | .9233 | .92224 |
| | | 2 | .8663 | .87104 |
| | | 3 | .7867 | .78752 |
| | | 4 | .7277 | .72960 |
| | | 5 | .6923 | .69120 |
| .40 | 10 | 1 | .9937 | .99395 |
| | | 2 | .9883 | .98614 |
| | | 3 | .9803 | .97906 |
| | | 4 | .9740 | .97392 |
| | | 5 | .9677 | .96953 |
| | | 6 | .9620 | .96331 |
| | | 7 | .9560 | .95596 |
| | | 8 | .9463 | .94895 |
| | | 9 | .9413 | .94365 |
| | | 10 | .9387 | .94101 |

## TABLE II

| SAMPLE SIZE (n) | CRITICAL VALUE | MASSEY | APPENDIX B |
|---|---|---|---|
| 5 | .20 | .0384 | .03840 |
| | .40 | .6521* | .69120 |
| | .60 | .9699 | .96992 |
| | .80 | .99935 | .99936 |
| 10 | .10 | .0004 | .00036 |
| | .20 | .2513 | .25128 |
| | .30 | .7331* | .72946 |
| | .40 | .9410 | .94101 |
| | .50 | .9922 | .99222 |
| | .60 | .9994 | .99943 |
| 20 | .10 | .0238 | .02374 |
| | .15 | .2955 | .29553 |
| | .20 | .6473 | .64728 |
| | .25 | .8624 | .86237 |
| | .30 | .9569 | .95693 |
| | .35 | .9892 | .98924 |
| | .40 | .9979 | .99787 |
| | .45 | .9997 | .99967 |
| 25 | .20 | .7637 | .76367 |
| | .24 | .9057 | .90564 |
| | .28 | .9683 | .96832 |
| | .32 | .9911 | .99109 |
| | .36 | .9979 | .99791 |
| 30 | .1667 | .6629 | .66289 |
| | .2000 | .8420 | .84202 |
| | .2333 | .9359 | .93587 |
| | .2667 | .9774 | .97744 |
| | .3000 | .9931 | .99314 |

*significance difference exists between Massey's results and results within Appendix B. Further study has indicated that the results of Appendix B are correct.

# VIII. APPLICATIONS OF MODIFIED KOLOMOGOROV-SMIRNOV TEST

The modified Kolomogorov-Smirnov test is essentially a Kolomogorov-Smirnov test run over only the first f ordered observations. Therefore, the modified test is applicable to data for which the Kolomogorov-Smirnov test would be applicable except that the data is censored at some predetermined number of ordered observations or, for some other reason, it is undesirable or impossible to consider datum beyond the specified number of observations. When the modified test uses all n ordered observations, the modified Kolomogorov-Smirnov test is identical to the Kolomogorov-Smirnov test.

As with the Kolomogorov-Smirnov test, the modified test can be used as a goodness of fit test if the (continuous) distribution function (F) is completely hypothesized. Furthermore, just as the probability statement for the Kolomogorov-Smirnov test can be inverted to yield a method of setting confidence limits (1, pg 457), the probability statement for the modified Kolomogorov-Smirnov test can be inverted to yield a method of setting confidence limits utilizing only the first f ordered sample observations. This attribute of the modified test can be useful in censored life testing situations where data is censored at some predetermined number of ordered observations.

## IX. WORK REMAINING TO BE DONE

Areas wherein it appears prudent to direct further investigation of the modified Kolomogorov-Smirnov test presented herein are:

(a) evaluation of the power of the modified test.

(b) comparison of the performance of the modified test against other test procedures presented in the literature.

(c) restructure the test to provide critical value versus significance level value for censoring at a predetermined point instead of a predetermined number of observations.

It is conjectured that item (c) can be accomplished by utilizing the difference equations presented by Massey [3] except carrying the summation over lattice points within the band only to the censor point and then translating the these values to the n/n lattice point by use of the OVUP(i,j) array presented in Section VI of this paper.

# X. SUMMARY

The Kolomogorov-Smirnov test is a well documented procedure for testing goodness of fit between a hypothesized distribution and an unknown distribution by utilizing a random sample drawn from the unknown distribution. The Kolomogorov-Smirnov test can be utilized to establish confidence limits for the hypothesized distribution function as a whole. The significance levels associated with given critical values are well tabulated and procedures for their calculations have been published in the literature.

A derivation and computational formulation for a modified Kolomogorov-Smirnov test has been presented. It extends the hypothesis testing and confidence limits advantages of the Kolomogorov-Smirnov test to one for censored data. The modified test is essentially a Kolomogorov-Smirnov test run only over the first f ordered sample observations. A listing of the computer program used in the calculations of significance levels of the modified test for given critical values is presented as Appendix A. The resulting significance levels for specified parameter values are given in Appendix B.

## APPENDIX A: COMPUTER PROGRAM FOR CALCULATION OF SIGNIFICANCE LEVELS

```
CALCULATION OF CENSORED PROBABILITIES, CV=K/N, K AN INTEGER
      DIMENSION Z(31,31), OVUP(30,30), FACT(31)
      DATA OVUP/900*0.0/, FACT/1.0,30*0.0/
      MIAC = 31
      DO 10 IAC=2,MIAC
      FACT(IAC) = FACT(IAC-1)*FLOAT(IAC-1)
   10 CONTINUE
      OVUP(1,1) = 1.0
      DO 40 I=2,30
      DO 30 J=1,30
      JAA = J+1
      OVUP(I,J) = 0.0
      DO 20 NH=1,J
      OVUP(I,J) = OVUP(I,J) + OVUP(I-1,NH)/FACT(J+1-NH)
   20 CONTINUE
   30 CONTINUE
   40 CONTINUE
VALUE IN OVUP(I,J) IS FOR OVER (I-1) AND UP (J-1)
THEREFORE, TO GO OVER I AND UP J CALL OVUP(I+1,J+1)
      DO 300 N=5,30,5
      WRITE (6,5) N
    5 FORMAT(5X,'N = ', I3)
      JJJ = N+1
      NO1 = MIN0(10,N)
      DO 200 K=1,NO1
      WRITE (6,6) K
    6 FORMAT(5X, 'K = ', I3)
      DO 110 NF=1,N
      DO 45 I=1,31
      DO 44 J=1,31
      Z(I,J) = 0.0
   44 CONTINUE
   45 CONTINUE
      Z(1,1) = 1.0
      MIA = MIN0(NF+K, N+1)
      DO 100 IA=2,MIA
      MINJA = MAX0(1, IA-K+1)
      IF (IA .EQ. N+1) MINJA=N+1
      MAXJA = N+1
      DO 90 JA=MINJA,MAXJA
      Z(IA,JA) = 0.0
      NHMIN = MAX0(1, IA-K)
      NHMAX = MIN0(K+IA-2, NF , JA)
      DO 70 NH=NHMIN,NHMAX
      Z(IA,JA) = Z(IA,JA) + Z(IA-1,NH)/FACT(JA-NH+1)
   70 CONTINUE
   90 CONTINUE
  100 CONTINUE
HAVE NOW COMPLETED CALCULATION OF THE Z AND OVUP VALUES FOR
PARAMETERS N,K,NF WHERE N=SAMPLE SIZE, K/N=CRITICAL VALUE
AND NF=# OBSERVATIONS OVER WHICH THE TEST IS RUN .
      IBMIN = MAX0(NF-K+2, 1)
      IBMAX = MIN0(NF+K, N+1)
      SUM = 0.0
      DO 150 IB=IBMIN,IBMAX
      JBMIN = NF+1
      IF (IB .EQ. N+1 ) JBMIN=N+1
      JBMAX = N+1
      DO 140 JB=JBMIN,JBMAX
      SUM = SUM + Z(IB,JB)*OVUP(N-IB+2, N-JB+2)
  140 CONTINUE
  150 CONTINUE
      PEN = SUM * FACT(N+1)/(FLOAT(N)**N)
      WRITE (6,600) NF, PEN
  600 FORMAT(9X,I3,' OBSERVATIONS - - - -',F9.5)
  110 CONTINUE
  200 CONTINUE
  300 CONTINUE
      STOP
```

# APPENDIX B: CALCULATED SIGNIFICANCE LEVELS

(n = sample size, f = number of observations, critical value = k/n)

n = 5

| f k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.67232 | 0.92224 | 0.98976 | 0.99967 | 0.99999 | | | | | |
| 2 | 0.28000 | 0.87104 | 0.98496 | 0.99968 | 1.00000 | | | | | |
| 3 | 0.12160 | 0.78752 | 0.98496 | 0.99968 | 1.00000 | | | | | |
| 4 | 0.05760 | 0.72960 | 0.97824 | 0.99968 | 1.00000 | | | | | |
| 5 | 0.03840 | 0.69120 | 0.96992 | 0.99936 | 1.00000 | | | | | |

n = 10

| f k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.65132 | 0.89262 | 0.97175 | 0.99395 | 0.99902 | 0.99989 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 2 | 0.25320 | 0.81192 | 0.94152 | 0.98614 | 0.99771 | 0.99977 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 3 | 0.09911 | 0.68126 | 0.91515 | 0.97906 | 0.99668 | 0.99972 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 4 | 0.03914 | 0.55947 | 0.88112 | 0.97392 | 0.99618 | 0.99971 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 5 | 0.01564 | 0.46168 | 0.83988 | 0.96953 | 0.99611 | 0.99971 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 6 | 0.00635 | 0.38556 | 0.80282 | 0.96331 | 0.99596 | 0.99971 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 7 | 0.00265 | 0.32827 | 0.77438 | 0.95596 | 0.99516 | 0.99970 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 8 | 0.00115 | 0.28842 | 0.75368 | 0.94895 | 0.99391 | 0.99963 | 0.99999 | 1.00000 | 1.00000 | 1.00000 |
| 9 | 0.00054 | 0.26563 | 0.73821 | 0.94368 | 0.99279 | 0.99951 | 0.99998 | 1.00000 | 1.00000 | 1.00000 |
| 10 | 0.00036 | 0.25128 | 0.72946 | 0.94101 | 0.99222 | 0.99943 | 0.99998 | 1.00000 | 1.00000 | 1.00000 |

n = 15

| f \ k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.64473 | 0.88310 | 0.96481 | 0.99045 | 0.99771 | 0.99952 | 0.99991 | 0.99998 | 0.99999 | 0.99999 |
| 2 | 0.24576 | 0.79514 | 0.92578 | 0.97675 | 0.99379 | 0.99862 | 0.99975 | 0.99996 | 0.99999 | 0.99999 |
| 3 | 0.09394 | 0.65514 | 0.88982 | 0.96212 | 0.98918 | 0.99750 | 0.99955 | 0.99993 | 0.99998 | 0.99999 |
| 4 | 0.03602 | 0.52460 | 0.84326 | 0.94813 | 0.98457 | 0.99640 | 0.99937 | 0.99992 | 0.99998 | 0.99999 |
| 5 | 0.01386 | 0.41916 | 0.78520 | 0.93326 | 0.98045 | 0.99549 | 0.99925 | 0.99991 | 0.99998 | 0.99999 |
| 6 | 0.00536 | 0.33562 | 0.72783 | 0.91412 | 0.97684 | 0.99486 | 0.99919 | 0.99991 | 0.99998 | 0.99999 |
| 7 | 0.00208 | 0.26985 | 0.67530 | 0.89308 | 0.97277 | 0.99450 | 0.99918 | 0.99991 | 0.99999 | 0.99999 |
| 8 | 0.00082 | 0.21822 | 0.62879 | 0.87270 | 0.96810 | 0.99413 | 0.99917 | 0.99991 | 0.99998 | 0.99999 |
| 9 | 0.00032 | 0.17779 | 0.58861 | 0.85455 | 0.96331 | 0.99347 | 0.99914 | 0.99991 | 0.99999 | 0.99999 |
| 10 | 0.00013 | 0.14631 | 0.55494 | 0.83946 | 0.95870 | 0.99250 | 0.99905 | 0.99991 | 0.99999 | 0.99999 |
| 11 | 0.00005 | 0.12206 | 0.52806 | 0.82755 | 0.95438 | 0.99140 | 0.99888 | 0.99990 | 0.99999 | 0.99999 |
| 12 | 0.00002 | 0.10388 | 0.50827 | 0.81793 | 0.95064 | 0.99035 | 0.99869 | 0.99988 | 0.99999 | 0.99999 |
| 13 | 0.00001 | 0.09125 | 0.49415 | 0.81038 | 0.94777 | 0.98953 | 0.99852 | 0.99986 | 0.99999 | 0.99999 |
| 14 | 0.00000 | 0.08404 | 0.48377 | 0.80523 | 0.94594 | 0.98902 | 0.99841 | 0.99984 | 0.99999 | 1.00000 |
| 15 | 0.00000 | 0.07950 | 0.47795 | 0.80275 | 0.94517 | 0.98882 | 0.99837 | 0.99983 | 0.99999 | 1.00000 |

n = 20

| f \ k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.64151 | 0.87842 | 0.96123 | 0.98846 | 0.99681 | 0.99919 | 0.99980 | 0.99994 | 0.99997 | 0.99998 |
| 2 | 0.24227 | 0.78722 | 0.91799 | 0.97155 | 0.99112 | 0.99751 | 0.99937 | 0.99985 | 0.99996 | 0.99998 |
| 3 | 0.09163 | 0.64328 | 0.87783 | 0.95298 | 0.98398 | 0.99520 | 0.99874 | 0.99971 | 0.99993 | 0.99997 |
| 4 | 0.03471 | 0.50956 | 0.82612 | 0.93455 | 0.97627 | 0.99253 | 0.99798 | 0.99953 | 0.99990 | 0.99997 |
| 5 | 0.01317 | 0.40200 | 0.76163 | 0.91448 | 0.96853 | 0.98975 | 0.99718 | 0.99935 | 0.99987 | 0.99996 |
| 6 | 0.00501 | 0.31713 | 0.69734 | 0.88859 | 0.96080 | 0.98708 | 0.99643 | 0.99919 | 0.99984 | 0.99996 |
| 7 | 0.00191 | 0.25053 | 0.63757 | 0.85935 | 0.95180 | 0.98462 | 0.99577 | 0.99906 | 0.99983 | 0.99996 |
| 8 | 0.00073 | 0.19836 | 0.58334 | 0.82944 | 0.94128 | 0.98215 | 0.99526 | 0.99898 | 0.99982 | 0.99996 |
| 9 | 0.00028 | 0.15748 | 0.53468 | 0.80048 | 0.92996 | 0.97940 | 0.99485 | 0.99894 | 0.99982 | 0.99996 |
| 10 | 0.00011 | 0.12544 | 0.49130 | 0.77337 | 0.91863 | 0.97641 | 0.99443 | 0.99892 | 0.99982 | 0.99996 |
| 11 | 0.00004 | 0.10032 | 0.45287 | 0.74859 | 0.90789 | 0.97333 | 0.99392 | 0.99888 | 0.99982 | 0.99996 |
| 12 | 0.00002 | 0.08063 | 0.41906 | 0.72646 | 0.89818 | 0.97033 | 0.99329 | 0.99880 | 0.99981 | 0.99996 |
| 13 | 0.00001 | 0.06519 | 0.38964 | 0.70720 | 0.88976 | 0.96751 | 0.99255 | 0.99867 | 0.99980 | 0.99996 |
| 14 | 0.00000 | 0.05311 | 0.36443 | 0.69100 | 0.88268 | 0.96488 | 0.99177 | 0.99850 | 0.99978 | 0.99996 |
| 15 | 0.00000 | 0.04370 | 0.34340 | 0.67795 | 0.87680 | 0.96251 | 0.99101 | 0.99833 | 0.99975 | 0.99996 |
| 16 | 0.00000 | 0.03646 | 0.32665 | 0.66783 | 0.87184 | 0.96049 | 0.99036 | 0.99816 | 0.99972 | 0.99996 |

| k<br>f | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 0.00000 | 0.03103 | 0.31434 | 0.65979 | 0.86785 | 0.95890 | 0.98984 | 0.99803 | 0.99970 | 0.99996 |
| 18 | 0.00000 | 0.02726 | 0.30558 | 0.65354 | 0.86493 | 0.95780 | 0.98949 | 0.99794 | 0.99968 | 0.99996 |
| 19 | 0.00000 | 0.02510 | 0.29914 | 0.64931 | 0.86312 | 0.95716 | 0.98930 | 0.99789 | 0.99967 | 0.99996 |
| 20 | 0.00000 | 0.02374 | 0.29553 | 0.64728 | 0.86237 | 0.95693 | 0.98924 | 0.99787 | 0.99967 | 0.99996 |

n = 25

| k<br>f | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.63960 | 0.87563 | 0.95906 | 0.98719 | 0.99620 | 0.99893 | 0.99970 | 0.99991 | 0.99996 | 0.99997 |
| 2 | 0.24023 | 0.78260 | 0.91337 | 0.96830 | 0.98931 | 0.99667 | 0.99903 | 0.99973 | 0.99991 | 0.99996 |
| 3 | 0.09031 | 0.63650 | 0.87087 | 0.94740 | 0.98052 | 0.99343 | 0.99798 | 0.99943 | 0.99984 | 0.99994 |
| 4 | 0.03399 | 0.50116 | 0.81639 | 0.92644 | 0.97079 | 0.98954 | 0.99663 | 0.99902 | 0.99973 | 0.99992 |
| 5 | 0.01280 | 0.39269 | 0.74856 | 0.90351 | 0.96075 | 0.98528 | 0.99507 | 0.99853 | 0.99960 | 0.99989 |
| 6 | 0.00483 | 0.30742 | 0.68088 | 0.87399 | 0.95043 | 0.98088 | 0.99342 | 0.99801 | 0.99946 | 0.99985 |
| 7 | 0.00182 | 0.24078 | 0.61785 | 0.84046 | 0.93829 | 0.97649 | 0.99177 | 0.99748 | 0.99933 | 0.99983 |
| 8 | 0.00069 | 0.18878 | 0.56049 | 0.80579 | 0.92397 | 0.97181 | 0.99018 | 0.99699 | 0.99920 | 0.99980 |
| 9 | 0.00026 | 0.14822 | 0.50876 | 0.77166 | 0.90820 | 0.96652 | 0.98864 | 0.99655 | 0.99910 | 0.99979 |
| 10 | 0.00010 | 0.11656 | 0.46231 | 0.73893 | 0.89181 | 0.96060 | 0.98704 | 0.99618 | 0.99903 | 0.99978 |

| f \ k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 0.00004 | 0.09184 | 0.42071 | 0.70806 | 0.87546 | 0.95428 | 0.98532 | 0.99584 | 0.99897 | 0.99977 |
| 12 | 0.00001 | 0.07252 | 0.38352 | 0.67924 | 0.85963 | 0.94785 | 0.98348 | 0.99550 | 0.99893 | 0.99977 |
| 13 | 0.00001 | 0.05740 | 0.35034 | 0.65256 | 0.84465 | 0.94155 | 0.98158 | 0.99513 | 0.99890 | 0.99977 |
| 14 | 0.00000 | 0.04557 | 0.32081 | 0.62808 | 0.83078 | 0.93560 | 0.97969 | 0.99470 | 0.99883 | 0.99976 |
| 15 | 0.00000 | 0.03630 | 0.29461 | 0.60582 | 0.81818 | 0.93014 | 0.97786 | 0.99423 | 0.99874 | 0.99975 |
| 16 | 0.00000 | 0.02903 | 0.27146 | 0.58582 | 0.80699 | 0.92528 | 0.97613 | 0.99373 | 0.99863 | 0.99974 |
| 17 | 0.00000 | 0.02333 | 0.25113 | 0.56813 | 0.79729 | 0.92104 | 0.97452 | 0.99320 | 0.99850 | 0.99971 |
| 18 | 0.00000 | 0.01886 | 0.23346 | 0.55283 | 0.78911 | 0.91738 | 0.97303 | 0.99270 | 0.99836 | 0.99969 |
| 19 | 0.00000 | 0.01537 | 0.21833 | 0.54000 | 0.78238 | 0.91423 | 0.97171 | 0.99224 | 0.99824 | 0.99966 |
| 20 | 0.00000 | 0.01265 | 0.20572 | 0.52969 | 0.77686 | 0.91153 | 0.97059 | 0.99184 | 0.99812 | 0.99964 |
| 21 | 0.00000 | 0.01055 | 0.19568 | 0.52170 | 0.77229 | 0.90933 | 0.96968 | 0.99153 | 0.99803 | 0.99962 |
| 22 | 0.00000 | 0.00898 | 0.18830 | 0.51536 | 0.76864 | 0.90766 | 0.96903 | 0.99131 | 0.99797 | 0.99960 |
| 23 | 0.00000 | 0.00789 | 0.18304 | 0.51045 | 0.76598 | 0.90652 | 0.96861 | 0.99118 | 0.99793 | 0.99959 |
| 24 | 0.00000 | 0.00726 | 0.17918 | 0.50713 | 0.76435 | 0.90587 | 0.96839 | 0.99111 | 0.99792 | 0.99959 |
| 25 | 0.00000 | 0.00687 | 0.17702 | 0.50553 | 0.76367 | 0.90564 | 0.96832 | 0.99109 | 0.99791 | 0.99959 |

n = 30

| f \ k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.63833 | 0.87378 | 0.95760 | 0.98633 | 0.99578 | 0.99875 | 0.99964 | 0.99990 | 0.99997 | 0.99998 |
| 2 | 0.23890 | 0.77958 | 0.91030 | 0.96611 | 0.98804 | 0.99605 | 0.99877 | 0.99964 | 0.99989 | 0.99996 |
| 3 | 0.08947 | 0.63210 | 0.86632 | 0.94367 | 0.97810 | 0.99212 | 0.99737 | 0.99919 | 0.99976 | 0.99993 |
| 4 | 0.03353 | 0.49579 | 0.81012 | 0.92108 | 0.96700 | 0.98732 | 0.99552 | 0.99854 | 0.99956 | 0.99987 |
| 5 | 0.01257 | 0.38683 | 0.74023 | 0.89636 | 0.95542 | 0.98196 | 0.99333 | 0.99775 | 0.99930 | 0.99980 |
| 6 | 0.00472 | 0.30143 | 0.67056 | 0.86460 | 0.94340 | 0.97630 | 0.99091 | 0.99683 | 0.99900 | 0.99971 |
| 7 | 0.00177 | 0.23488 | 0.60571 | 0.82850 | 0.92924 | 0.97049 | 0.98836 | 0.99585 | 0.99867 | 0.99961 |
| 8 | 0.00067 | 0.18312 | 0.54670 | 0.79104 | 0.91250 | 0.96420 | 0.98578 | 0.99484 | 0.99832 | 0.99951 |
| 9 | 0.00025 | 0.14289 | 0.49348 | 0.75399 | 0.89392 | 0.95700 | 0.98313 | 0.99384 | 0.99799 | 0.99942 |
| 10 | 0.00009 | 0.11160 | 0.44566 | 0.71825 | 0.87440 | 0.94886 | 0.98029 | 0.99286 | 0.99767 | 0.99933 |
| 11 | 0.00004 | 0.08726 | 0.40277 | 0.68424 | 0.85461 | 0.94001 | 0.97717 | 0.99190 | 0.99738 | 0.99925 |
| 12 | 0.00001 | 0.06831 | 0.36436 | 0.65215 | 0.83503 | 0.93074 | 0.97376 | 0.99089 | 0.99712 | 0.99919 |
| 13 | 0.00001 | 0.05355 | 0.32997 | 0.62203 | 0.81602 | 0.92134 | 0.97015 | 0.98983 | 0.99688 | 0.99915 |
| 14 | 0.00000 | 0.04204 | 0.29922 | 0.59388 | 0.79779 | 0.91203 | 0.96644 | 0.98871 | 0.99663 | 0.99911 |
| 15 | 0.00000 | 0.03306 | 0.27173 | 0.56766 | 0.78049 | 0.90301 | 0.96273 | 0.98755 | 0.99638 | 0.99907 |

| f \ k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 0.00000 | 0.02605 | 0.24717 | 0.54332 | 0.76424 | 0.89444 | 0.95913 | 0.98638 | 0.99609 | 0.99903 |
| 17 | 0.00000 | 0.02057 | 0.22526 | 0.52081 | 0.74912 | 0.88643 | 0.95572 | 0.98523 | 0.99579 | 0.99897 |
| 18 | 0.00000 | 0.01628 | 0.20574 | 0.50011 | 0.73519 | 0.87908 | 0.95256 | 0.98411 | 0.99547 | 0.99889 |
| 19 | 0.00000 | 0.01292 | 0.18837 | 0.48118 | 0.72252 | 0.87246 | 0.94968 | 0.98305 | 0.99514 | 0.99881 |
| 20 | 0.00000 | 0.01029 | 0.17297 | 0.46400 | 0.71115 | 0.86659 | 0.94711 | 0.98205 | 0.99480 | 0.99871 |
| 21 | 0.00000 | 0.00823 | 0.15937 | 0.44860 | 0.70114 | 0.86150 | 0.94483 | 0.98112 | 0.99447 | 0.99861 |
| 22 | 0.00000 | 0.00662 | 0.14743 | 0.43499 | 0.69251 | 0.85715 | 0.94282 | 0.98027 | 0.99415 | 0.99852 |
| 23 | 0.00000 | 0.00535 | 0.13706 | 0.42323 | 0.68527 | 0.85347 | 0.94105 | 0.97952 | 0.99388 | 0.99843 |
| 24 | 0.00000 | 0.00436 | 0.12817 | 0.41338 | 0.67932 | 0.85033 | 0.93954 | 0.97888 | 0.99364 | 0.99835 |
| 25 | 0.00000 | 0.00359 | 0.12077 | 0.40546 | 0.67446 | 0.84769 | 0.93829 | 0.97836 | 0.99345 | 0.99830 |
| 26 | 0.00000 | 0.00299 | 0.11487 | 0.39933 | 0.67043 | 0.84556 | 0.93732 | 0.97797 | 0.99331 | 0.99825 |
| 27 | 0.00000 | 0.00255 | 0.11054 | 0.39446 | 0.66723 | 0.84395 | 0.93662 | 0.97770 | 0.99322 | 0.99822 |
| 28 | 0.00000 | 0.00224 | 0.10746 | 0.39070 | 0.66491 | 0.84286 | 0.93617 | 0.97754 | 0.99317 | 0.99821 |
| 29 | 0.00000 | 0.00206 | 0.10519 | 0.38815 | 0.66348 | 0.84224 | 0.93594 | 0.97746 | 0.99314 | 0.99820 |
| 30 | 0.00000 | 0.00195 | 0.10392 | 0.38693 | 0.66289 | 0.84202 | 0.93587 | 0.97744 | 0.99314 | 0.99820 |

# APPENDIX C: COMPUTER PROGRAM FOR VERIFICATION OF CALCULATED SIGNIFICANCE LEVELS

```
ICNT(I) IS THE NUMBER OF REPLICATIONS FOR WHICH THE FIRST I
ORDERED OBSERVATIONS REMAIN WITHIN THE BAND  WHERE F IS
THE U(0,1) DISTRIBUTION, CV=K/N IS THE CRITICAL VALUE, AND
N IS THE SAMPLE SIZE.
      DIMENSION X(30),T(30),ICNT(30)
      DATA ICNT/30*0/
      MR = 12493
      IR = 87345
      KR = 8*MR + 3
      NREP = 3000
      DO 550 N=5,10,5
      K = 2*N/5
      CV = FLOAT(K)/FLOAT(N)
      WRITE (6,50) CV
   50 FORMAT (3X,'CRIT VALUE =',F6.4)
      DO 500 KKK=1,NREP
      DO 210 K=1,N
      IR = IR*KR
      R = 0.5 + FLOAT(IR)*2.328306E-10
      X(K) = R
  210 CONTINUE
      DO 240 L=1,N
      T(L) = 1.01
      DO 230 K=1,N
      IF (X(K) .GT. T(L) ) GO TO 230
      KO=K
      T(L) = X(K)
  230 CONTINUE
      X(KO) = 1.01
  240 CONTINUE
      DO 300 I=1,N
      DN = T(I) - FLOAT(I-1)/FLOAT(N)
      DN2 = FLOAT(I)/FLOAT(N) - T(I)
      IF (DN2 .GT. DN ) DN = DN2
      IF (DN .GT. CV ) GO TO 500
      ICNT(I) = ICNT(I) + 1
  300 CONTINUE
  500 CONTINUE
      DO 550 K=1,N
      PP = FLOAT(ICNT(K))/FLOAT(NREP)
      WRITE (6,600) K,PP
  600 FORMAT (5X,I3,F10.4)
      ICNT(K) = 0
  550 CONTINUE
      STOP
```

# BIBLIOGRAPHY

1. Maurice G. KENDALL and Alan STUART, The Advanced Theory of Statistics, VOL 2, Hafner Publishing Company, New York (1961) pg 452 - 461.

2. Frank J. MASSEY, Jr., The Kolomogorov-Smirnov Test for Goodness of Fit, American Statistical Association Journal, 46, 1951, pg 68 - 77.

3. Frank J. MASSEY, Jr., A Note on the Estimation of a Distribution Function by Confidence Limits, Annals of Mathematical Statistics, 21, 1950, pg 116 - 119.

4. Leslie H. MILLER, Table of Percentage Points of Kolomogorov Statistics, American Statistical Association Journal, 51, 1956, pg 111 - 121.

# INITIAL DISTRIBUTION LIST

| | | No. Copies |
|---|---|---|
| 1. | Defense Documentation Center<br>Cameron Station<br>Alexandria, Virginia 22314 | 2 |
| 2. | Library, Code 0212<br>Naval Postgraduate School<br>Monterey, California 93940 | 2 |
| 3. | Assoc Professor D. R. Barr, Code 55Bn<br>Department of Operations Research &<br>Administrative Sciences<br>Naval Postgraduate School<br>Monterey, California 93940 | 1 |
| 4. | Chief of Naval Personnel<br>PERS - 11b<br>Department of the Navy<br>Washington, D. C. 20370 | 1 |
| 5. | Naval Postgraduate School<br>Department of Operations Research &<br>Administrative Sciences<br>Monterey, California 93940 | 1 |
| 6. | LT Teddy G. Davidson, USN<br>Rt 1, Box 86<br>Hazen, North Dakota 58545 | 1 |

## DOCUMENT CONTROL DATA - R & D

*(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)*

| | |
|---|---|
| 1. ORIGINATING ACTIVITY *(Corporate author)*<br>Naval Postgraduate School<br>Monterey, California 93940 | 2a. REPORT SECURITY CLASSIFICATION<br>UNCLASSIFIED<br>2b. GROUP |

3. REPORT TITLE
A MODIFIED KOLOMOGOROV-SMIRNOV TEST APPLICABLE TO CENSORED SAMPLES

4. DESCRIPTIVE NOTES *(Type of report and, inclusive dates)*
Master's Thesis; September 1971

5. AUTHOR(S) *(First name, middle initial, last name)*
Teddy George Davidson

| | | |
|---|---|---|
| 6. REPORT DATE<br>September 1971 | 7a. TOTAL NO. OF PAGES<br>44 | 7b. NO. OF REFS<br>4 |

| | |
|---|---|
| 8a. CONTRACT OR GRANT NO | 9a. ORIGINATOR'S REPORT NUMBER(S) |
| b. PROJECT NO. | |
| c. | 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* |
| d. | |

10. DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited.

| | |
|---|---|
| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY<br>Naval Postgraduate School<br>Monterey, California 93940 |

13. ABSTRACT

Herein is presented a derivation and computational formulation for a modified Kolomogorov-Smirnov test. This test extends the hypothesis testing and confidence limits advantages of the Kolomogorov-Smirnov test to data which is censored beyond a predetermined number of observations.

A listing of the computer program used in the calculations of significance levels and the resulting significance levels for specified parameter values for the modified test are included.

| KEY WORDS | LINK A |  | LINK B |  | LINK C |  |
|---|---|---|---|---|---|---|
|  | ROLE | WT | ROLE | WT | ROLE | WT |
| Censored Testing |  |  |  |  |  |  |
| Hypothesis Testing |  |  |  |  |  |  |
| Kolomogorov-Smirnov Test |  |  |  |  |  |  |

Thesi
D166
c.1

Thesis 130752
D166 Davidson
c.1
A modified Kolomogorov-Smirnov test
applicable to censored
samples.

22 APR 83
BINDERY
27846

Thesis 130752
D166 Davidson
c.1
A modified Kolomogorov-Smirnov test
applicable to censored
samples.

thesD166
A modified Kolomogorov-Smirnov test appl
3 2768 001 02324 5
DUDLEY KNOX LIBRARY